Typology of Risks of Generative Text-to-Image Models

Year: 2023 | Citations: 109 | Authors: Charlotte M. Bird, Eddie L. Ungless, Atoosa Kasirzadeh

Abstract

This paper investigates the direct risks and harms associated with modern text-to-image generative models, such as DALL-E and Midjourney, through a comprehensive literature review. While these models offer unprecedented capabilities for generating images, their development and use introduce new types of risk that require careful consideration. Our review reveals significant knowledge gaps concerning the understanding and treatment of these risks despite some already being addressed. We offer a taxonomy of risks across six key stakeholder groups, inclusive of unexplored issues, and suggest future research directions. We identify 22 distinct risk types, spanning issues from data bias to malicious use. The investigation presented here is intended to enhance the ongoing discourse on responsible model development and deployment. By highlighting previously overlooked risks and gaps, it aims to shape subsequent research and governance initiatives, guiding them toward the responsible, secure, and ethically conscious evolution of text-to-image models.